

# Virtual Reality Simulation for the Operating Room

## *Proficiency-Based Training as a Paradigm Shift in Surgical Skills Training*

Anthony G. Gallagher, PhD,\* E. Matt Ritter, MD,† Howard Champion, MD,† Gerald Higgins, PhD,§ Marvin P. Fried, MD,‡ Gerald Moses, PhD,|| C. Daniel Smith, MD,\* and Richard M. Satava, MD¶

**Summary Background Data:** To inform surgeons about the practical issues to be considered for successful integration of virtual reality simulation into a surgical training program. The learning and practice of minimally invasive surgery (MIS) makes unique demands on surgical training programs. A decade ago Satava proposed virtual reality (VR) surgical simulation as a solution for this problem. Only recently have robust scientific studies supported that vision

**Methods:** A review of the surgical education, human-factor, and psychology literature to identify important factors which will impinge on the successful integration of VR training into a surgical training program.

**Results:** VR is more likely to be successful if it is systematically integrated into a well-thought-out education and training program which objectively assesses technical skills improvement proximate to the learning experience. Validated performance metrics should be relevant to the surgical task being trained but in general will require trainees to reach an objectively determined proficiency criterion, based on tightly defined metrics and perform at this level consistently. VR training is more likely to be successful if the training schedule takes place on an interval basis rather than massed into a short period of extensive practice. High-fidelity VR simulations will confer the greatest skills transfer to the in vivo surgical situation, but less expensive VR trainers will also lead to considerably improved skills generalizations.

**Conclusions:** VR for improved performance of MIS is now a reality. However, VR is only a training tool that must be thoughtfully introduced into a surgical training curriculum for it to successfully improve surgical technical skills.

(*Ann Surg* 2005;241: 364–372)

Simulation for the development and refinement of surgical skills has come to the forefront in recent years. There are a number of reasons for this. The rapid expansion of minimally invasive surgery (MIS) has demonstrated that the traditional model of “see one, do one, teach one” is not an optimal approach for training surgical skills. The MIS revolution has forced the surgical community to rethink how they train residents and adapt to new technologies. This rethinking has driven the search for innovative training solutions. Over a decade ago, Satava<sup>1</sup> proposed early adoption of virtual reality (VR) as a training tool. Despite this early vision, VR simulation has only recently begun to be accepted by the surgical community.<sup>2</sup> While the reasons for this delayed acceptance are numerous,<sup>3</sup> the initial lack of robust scientific evidence to support the use of VR for skills training and the lack of knowledge of how to effectively apply simulation to a surgical skills training program are the 2 most likely reasons for delayed adoption of this technology. It took a multidisciplinary team who looked beyond the “prettiness” of the simulator and drew from more than 50 years of sound research on aviation simulation and 100 years of behavioral sciences research to demonstrate the power of simulation for surgical training.<sup>4</sup> The remainder of this paper will outline how to correctly apply VR and simulation technology to a training program and identify key factors that will help assure successful implementation. While most of the examples are from MIS, they apply to most areas of medical training.

### Simulators: Education or Training

There seems to be some confusion as to whether simulators educate or train individuals, and the 2 terms are often

From the \*Emory Endosurgery Unit, Department of Surgery, Emory University School of Medicine, Atlanta, Georgia; †NCA Medical Simulation Center, Norman M. Rich Department of Surgery, Uniformed Services University of Health Sciences, Bethesda, Maryland; ‡Department of Otolaryngology, Albert Einstein College of Medicine and Montefiore Medical Center, Bronx, New York; §Federation of American Scientists, Washington, DC; ||Telemedicine and Advanced Technology Research Center (TATRC), Fort Detrick, Maryland; and the ¶Department of Surgery, University of Washington Medical Center, Seattle, Washington. Reprints: Anthony G. Gallagher, PhD, Director of Research, Emory Endosurgery Unit, Emory University School of Medicine, 1364 Clifton Road NE, Suite H-122, Atlanta, GA 30322. E-mail [aggallagher@comcast.net](mailto:aggallagher@comcast.net).

Copyright © 2005 by Lippincott Williams & Wilkins  
ISSN: 0003-4932/05/24102-0364  
DOI: 10.1097/01.sla.0000151982.85062.80

used interchangeably. Simulation is frequently referred to as “education” rather than “training” or education and training. Although closely related, education and training are not the same. Education usually refers to the communication or acquisition of knowledge or information, while training refers to the acquisition of skills (cognitive or psychomotor). Most currently available simulators primarily provide training. Furthermore, simulators alone provide only part of a curriculum. Individuals being prepared to perform a specific procedure need to know what to do, what not to do, how to do what they need to do, and how to identify when they have made a mistake. The trainer needs to know how the trainee is progressing and/or where they are on their learning curve, not just psychomotor learning but cognitive. Simulators are only part of the training solution confronting residency programs and credentialing committees around the world. Their power can only truly be realized if they are integrated into a well-thought-out curriculum.

A proposed template for developing a curriculum should include the following sequence: (1) didactic teaching of relevant knowledge (ie, anatomy, pathology, physiology); (2) instruction on the steps of the task or procedure; (3) defining and illustrating common errors; (4) test of all previous didactic information to insure the student understands all the cognitive skills before going to the technical skills training and in particular to be able to determine when they make an error; (5) technical skills training on the simulator; (6) provide immediate (proximate) feedback when an error occurs; (7) provide summative (terminal) feedback at the completion of a trial; (8) iterate the skills training (repeated trials) while providing evidence at the end of each trial of progress (graphing the “learning curve”), with reference to a proficiency performance goal that the trainee is expected to attain. While the above is a proposed template, it includes in a stepwise fashion all components published in the literature that would comprise a comprehensive and validated training curriculum. Details of how to create and implement such a curriculum will be discussed below.

### Skills Transfer and Skills Generalization

Before delving into the nuts and bolts of applying simulation, the concepts of skills transfer and skills generalization deserve some brief attention. Although these 2 phenomena are related and both refer to the process of skill acquisition, they are fundamentally different. *Skills generalization* refers to the training situation where the trainee learns fundamental skills that are crucial to completion of the actual operative task or procedure. *Skills transfer* refers to a training modality that directly emulates the task to be performed in vivo or in the testing condition. A practical example of the difference between skills generalization and transfer can be taken from the game of golf. Every golf pro will have beginning golfers practice swinging without even holding a

club. This would be skills generalization. The swing is crucial to executing any golf shot, but swinging without a club does not directly relate to a shot. An example of skills transfer would be a golfer repeatedly hitting a sand wedge out of the right side trap near the 18th green. If during the next round the golfer finds himself in that trap, the practiced skills will directly transfer to his current situation. In the world of simulation, a good example of skills generalization is the Minimally Invasive Surgical Trainer–Virtual Reality (MIST VR; Mentice, Gothenburg, Sweden) laparoscopic surgical training system. This system teaches basic psychomotor skills fundamental to performing a safe laparoscopic cholecystectomy, as well as many skills required in advanced laparoscopic procedures. The VR tasks do not resemble the operative field, but it has been clearly demonstrated that subjects who trained on the MIST VR performed a 2-handed laparoscopic cholecystectomy faster with fewer intraoperative errors.<sup>4</sup> It has also been demonstrated that these skills improve laparoscopic intracorporeal suturing.<sup>5</sup> These are 2 good examples of skills generalization, which represents a very powerful but misunderstood learning and training methodology. Simulators which rely on skills transfer might include manikin-type simulators such as Virgil (CIMIT, Boston, MA), high-end VR simulators such as both the Lap Mentor and GI Mentor II (Symbionix USA, Cleveland, OH), the Vascular Intervention System Training (VIST) (Mentice), and the ES3 (Lockheed Martin, Bethesda, MD). The simulated procedures look and feel similar to the actual procedures and will train skills that will directly transfer to the performed procedures.

A common mistake made by many trainers is that only simulators that provide a high-fidelity experience improve performance. This is inaccurate, as clearly demonstrated by the Yale VR to OR study mentioned above. The question that should be asked is, does the simulator train the appropriate skill to perform the procedure? It should also be noted that as fidelity increases, so does price. One of the most sophisticated VR simulators in the world is the VIST system, which simulates in real time, a full physics model of the vascular system. However, it costs \$300,000 per unit. Not all training programs can afford this level of simulation. Trainers must look beyond face validity of simulators and ask more important questions, such as, does it work (ie, train the appropriate skills), how well does it work, and how good is the evidence? This may involve trainers developing realistic expectations of what simulators should look like, which in turn will involve a genuine understanding of what simulations should be capable of achieving in a training program.

### Simulation Applied

#### Cognition

There are many aspects of human behavior that are crucial for safe MIS performance, and none seems to be less

understood by surgeons than cognition.<sup>6</sup> From reading the surgical literature, most surgical researchers appear to interpret cognition as knowledge or decision making. Undoubtedly, knowledge is crucial for the practice of surgery, and surgeons have done an exemplary job in the knowledge education of medical students and residents. However, behavioral or cognitive scientists would have some difficulties with the very restricted interpretation of knowledge or decision making as cognition. Knowledge and decision making are really only the end product of a process of perception, attention, information processing, information storage (including organization), and then retrieval from long-term memory at the appropriate time as “knowledge” which helps the individual make their decision.<sup>7</sup> One aspect of cognition that has received no consideration in the surgery literature is “attention,” but it is of paramount importance to the surgeon learning a new task or set of skills.

Attention usually refers to the ability to focus mental powers upon an object such as careful observing or listening, or the ability to concentrate mentally. It has been known for at least half a century that human beings have a limited attentional capacity.<sup>8</sup> This means that we can only attend to a finite amount of information or stimuli at any given time. Figure 1 shows a diagrammatic representation of attentional resources used by a master surgeon and a novice surgeon for different aspects of operative performance. The master surgeon occupies less attentional capacity than the novice for

basic psychomotor, spatial, and decision-making processes than does the novice. The space between the top of the master surgeon column and the attentional capacity threshold represents an attentional resource buffer zone that the master surgeon uses to monitor for and manage complications and keep track of additional data such as instrument readouts or information from the patient’s physiologic monitors. When a novice is acquiring new skills such as those required for laparoscopic surgery, he or she must use these attentional resources to consciously monitor what their hands are doing in addition to the spatial judgments and operative decision making, even though their amount of decision making is somewhat less than the attending master surgeon. This results in limited additional attentional capacity for the novice. Given that most novice surgeons are supposed to be learning the judgment and decision making from the master surgeon, these additional attentional resource requirements quickly exceed the novice’s attentional threshold. What simulation skills training allows for is the development of the “pretrained novice” (Fig. 2). This individual has been trained using simulation to the point where many of the psychomotor skills and spatial judgments have been automated, meaning that they now occupy significantly fewer attentional resources, allowing the novice to focus more on learning the steps of the operation and how to handle complications instead of wasting valuable OR time on the initial refinement of technical skills.

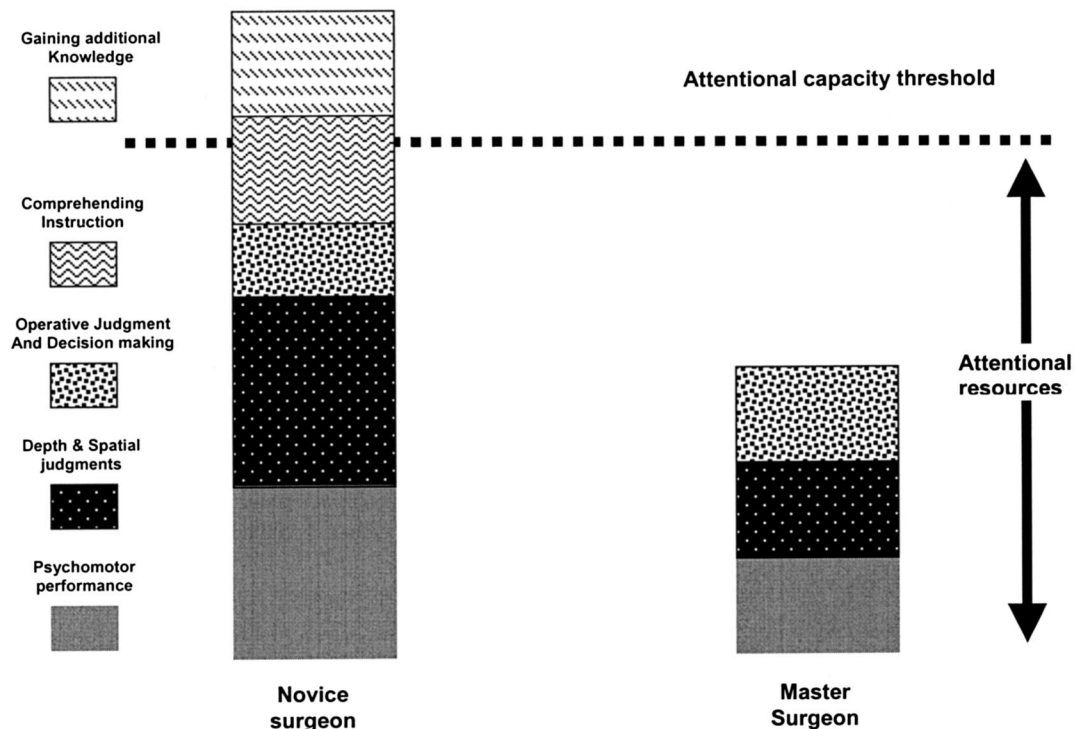


FIGURE 1. Hypothetical model of attention.

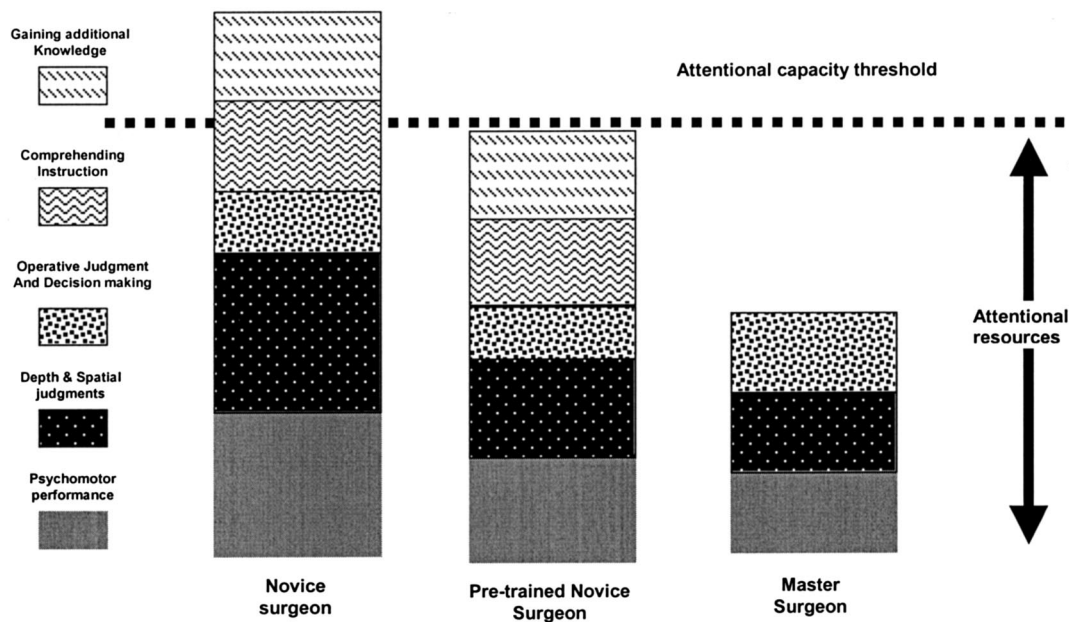


FIGURE 2. Hypothetical attentional resource benefits of simulation training.

This automation process is represented in Figure 3. There are 2 major factors that determine the rate of automation. One is the level of fundamental mental abilities that surgeon brings to the table, and the other is the experience they have gained. The more innate visuospatial, perceptual, and psychomotor ability the surgeon has, the faster he or she will automate the surgical skills, thus requiring fewer attentional resources to monitor basic aspects of their performance.<sup>9,10</sup> Figure 3 should be familiar to most surgeons as it appears to resemble “learning curves” associated with skill acquisition. Traditionally, learning curves have been reported just as a relationship between some proxy for skills automation such as procedure time and a surgeon’s experience expressed as the number of cases performed. However, as many surgeon educators are all too aware, the number of

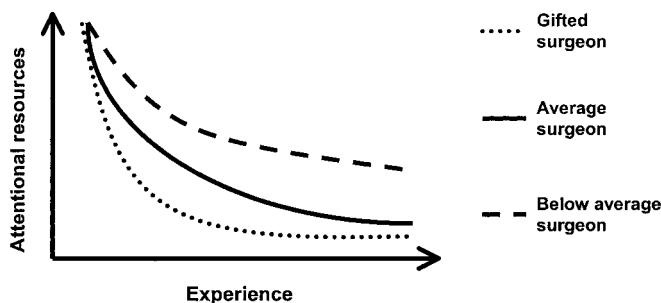


FIGURE 3. Hypothetical model of attentional resources used as a function of experience. Rate of reduction of attentional resource utilization is dependent upon the fundamental abilities of the surgeon.

procedures performed by a learner is at best a crude predictor of actual operative performance. As illustrated in Figure 3, the rate of automation will vary widely among surgeons based on their fundamental abilities. Given this variation, it would be nearly impossible to establish an experienced-based (ie, number of cases or number of repetitions on a simulator) criterion that would correspond to skills automation. A better predictor is objective assessment of technical skills performance.

The goal of any surgical training program should be to help the junior surgeon automate these basic psychomotor skills before they operate on a patient. This is one of the major advantages of simulation; it allows the trainee to automate in a risk-free environment and the trainer to monitor the automation process. Establishing when automation has been achieved will be dealt with later in this paper under performance levels.

The one aspect of cognition that surgery has excelled at is knowledge communication and assessment. However, knowledge alone does not make a surgeon. Rather, a surgeon is an individual who knows what to do, when to do it, and has the psychomotor skills to do it. Gallagher and Satava<sup>11</sup> showed that junior (>10 but <100 cases) and novice (no previous experience) laparoscopic subjects perform as well as experienced (>100 cases) laparoscopic surgeons after about 3 trials on moderately difficult VR laparoscopic psychomotor tasks. However, they also found that both the junior and experienced laparoscopic surgeons performed significantly better on a task utilizing electrocautery. Although the junior laparoscopic surgeons performed similar to the novices on

trials 1 and 2, their performance approached the experienced surgeons by trial 3. The novices continued to demonstrate poor performance across all 10 trials. The authors concluded that the reason for this was that the novices did not know that overuse of electrocautery when operating can be dangerous. They did not have the context upon which to base their skills training. Experienced surgeons knew that excess electrocautery can be dangerous, and the junior laparoscopic surgeons had probably been taught this but had forgotten it until they were reminded by the alarm associated with misuse or overuse of electrocautery on the VR task. The junior surgeons subsequently modified their behavior to be more like that of the experienced laparoscopic surgeons. The important point is that for optimal performance both knowledge and psychomotor skills must be acquired together, and, as well as being taught what to do, trainees should also be taught what not to do.

### Skill Acquisition

Psychomotor skill acquisition is an essential prerequisite for safe surgery. Traditionally, in surgery skills have been acquired by trainees through an apprenticeship model. Trainees observe the senior surgeons and performed under their direct guidance until “mastery” was achieved. Due to basic nature of open surgical skills, this model served surgical training well for more than a century.

In the mid to late 1980s, the advent and widespread demand for laparoscopic cholecystectomy disrupted the fundamentals of traditional surgical training. It was not only residents who needed training but now practicing surgeons who did not have the luxury of time to acquire these new and deceptively difficult skills. This MIS revolution eventually led surgeons to rethink how best to train surgical skills.

One approach to skills acquisition was use of simulation, proposed by Satava in the early 1990s. While surgical skills simulators are being produced in ever-increasing numbers, there is still confusion about how to use simulators to teach surgical skills. There seems to be a prescriptive approach to use of simulators in training. Trainers require trainees to “train” on simulators without much systematic thought about what they are trying to achieve. The underlying assumption seems to be that individuals who have performed the required number of procedures will be a safe practitioner. A fundamental flaw with this approach, as we have highlighted above, is that it ignores individual variability with respect to skill acquisition. Setting a fixed number of procedures or number of training hours is not an optimal approach to learning. In the next several sections, we will discuss the optimal application of simulation for skills acquisition, specifically addressing pretraining education, task configuration, training schedule, metrics, and feedback

### Pretraining Education

An optimal training strategy for any skill acquisition program would first ensure that the subject had sufficient knowledge of what to do, why to do it, and when and where to do it. It is also important that they know what *not* to do. This can be achieved in any number of ways, including books, journal publication, videos, lectures, and interactive seminars. However, independent of the method of delivery, an optimal training program would objectively check that the trainee does in fact know what he or she is supposed to know prior to beginning training. This objectively assessed pretraining education not only ensures that the trainee knows both what to and not to do but also allows the trainee to possess an educational context which will help him or her understand the skills training objectives.

### Task Configuration and Training Strategies

Assuming that the trainee has successfully acquired the appropriate level of knowledge, he or she should then begin a psychomotor skill acquisition program which “shapes” the appropriate performance in the correct sequence in a skills laboratory. This is known in the behavioral science literature as “shaping.” The term simply means that successive approximations of the desired response pattern are reinforced until the desired response occurs. For example, it would be inappropriate to expect a surgeon with no prior laparoscopic experience to come into the skills laboratory and start intracorporeal suturing immediately. This situation should be tackled by starting the surgeon off on relatively simple hand-eye coordination tasks that gradually become more difficult.<sup>12</sup> The transition in degree of difficulty should be smooth and relatively effortless for the subject with feedback proximate to performance. Proximate feedback leads to optimal learning. Many of the simulators that currently exist for training laparoscopic skills do indeed use shaping as their core training methodology. Tasks are configurable from easy, medium, and difficult settings, and tasks can be ordered so that they become progressively more difficult. However, it is not clear whether the software engineers were aware that this was what they were doing when they wrote the software. Also, this is only one training strategy that could be used.

Another training strategy is “fading” and is used by a number of simulators such as the GI Mentor II (Symbionix USA) and Endoscopic Sinus Simulator (ES3; Lockheed Martin). This strategy involves giving trainees major clues and guides at the start of training. Indeed, trainees might even begin with abstract tasks that elicit the same psychomotor performance as would be required to perform the task in vivo. As tasks become gradually more difficult, the amount of clues and guides is gradually faded out until the trainee is required to perform the task without support. For example, the ES3 simulator on easy level requires the trainee to navigate an instrument through a series of hoops, the path of which

mirrors the nasal cavity. The abstract task teaches the trainee the optimal path without having to worry about anatomic structure. The intermediate level requires the trainee to perform the same task; however, on this setting the hoops are overlaid on simulated nasal cavity tissue and anatomic landmarks. The third and most difficult level gives no aids. This has in effect been faded out.

Another very effective training strategy known as “backward chaining”<sup>13</sup> does not appear to have been used by any of the simulation companies. While shaping starts at the very beginning or basic steps of a skill or psychomotor task and gradually increases the complexity of the task requirement, backward chaining starts at the opposite end of the task (ie, the complete task minus 1 step). This training strategy was developed for tasks that are particularly difficult and frustrating to learn. A good example of a task that would fit this category is intracorporeal suturing and knot tying. An example of how this would be achieved is to break the procedure down into discrete psychomotor performance units (task deconstruction).

A number of researchers have done this for their teaching curriculum but then proceeded to require trainees to perform the complete task.<sup>14</sup> The problem with this approach is that the trainee has a high failure-to-success ratio, resulting in frustration, which in turn usually means that they give up trying to learn suturing. Backward chaining specifically programs a high success-to-failure ratio, thus reducing or eliminating learner frustration. Using the example of tying a laparoscopic intracorporeal slip-square knot, tasks would be set up so that the trainee does the last step first, ie, tying the final square knot. Trainees would continue to do this until they could do it proficiently every time. The next training task would involve trainees cinching or sliding the knot down into place and then squaring it off. Both steps would continue to be practiced until they are being performed consistently. The next training task would involve the square knot to a slip knot and then the 2 previous steps. This process would continue all the way back to the first step, ie, the formation of a “C” loop and the wrap and so on. The beauty of this approach to training is that only 1 new step is being added with each backward step or “chain” and that the forward chained behaviors had already been mastered, ensuring a high level of task success and a low level of frustration. In the box-trainer environment, this approach would have been very time consuming for the trainer preparing the backward-chained task configuration, as well as difficult to assess. These difficulties disappear in virtual space. Furthermore, at least 3 VR companies currently have suturing tasks that could be configured this way (ie, Mentice; SimSurger, Norway, Surgical Science, Sweden).

A final point to make on task configuration concerns what the task actually teaches. An implicit assumption here has been that simulators teach “good” behaviors, which is

ideally what they should do, but sometimes this is not always the case. Simulators can also teach “bad” behaviors such as putting clips on at the wrong angle without penalty or continuous use of electrocautery. This issue is important because it is very easy to teach bad behaviors and very difficult to eliminate them. Considerable care should be taken not to teach them in the first place. This will require experts (surgeons and trainers) evaluating any new VR tasks that are developed for this purpose. Also, many surgical tasks can be performed in different ways, reflecting a particular surgeon’s preference. Training to only 1 method has distinct limitations. Simulator designers should be flexible enough to allow for some variation in performance methods as long as they are safe and achieve the desired outcome.

### Training Schedule

Once the education and tasks are in place, trainees must practice the task to gain the skills. Extensive research has been conducted to determine the effects of practice schedules on the acquisition of simple motor skills.<sup>13</sup> Among the possible variables affecting the acquisition of motor skills, none has been more extensively studied than practice regimen.

Distribution of practice refers to the spacing of practice sessions either in one long session (massed practice) or multiple, short practice sessions (interval practice). Metalis investigated the effects of massed versus interval practice on the acquisition of video-game-playing skill.<sup>15</sup> Both the massed and interval practice groups showed marked improvement; however, the interval practice group consistently showed more improvement. Studies conducted in the 1940s and 1950s attempted to address the effects of massed as compared with interval practice. The majority of these studies showed that interval practice was more beneficial than massed practice. At present, most new skills are taught in massed sessions often lasting 1 or 2 days. The surgeons are often considered “trained” in this new technology after the short course, and the issues of competence and supervision of the newly trained surgeons are relegated to the individual hospital.

Why is interval practice a more effective training schedule than massed practice? A likely explanation is that the skills being learned have more time to be cognitively consolidated between practices. Consolidation is the process that is assumed to take place after acquisition of a new behavior. The process assumes long-term neurophysiological changes that allow for the relatively permanent retention of learned behavior. Scientific evidence for this process is now starting to emerge.<sup>16</sup> While the underlying mechanisms may be academically interesting, practically speaking it is clear that distributed interval training is superior to massed training, and thus, this type of training schedule should be incorporated into any optimally designed curriculum.

## Establishing Proficiency Criterion

How long should someone train on a simulator or how many times should they complete a given task? This question is akin to the question, how long is a string? The answer depends on the individual trainee or string. Trainees begin with different levels of fundamental abilities, skills, and motivation. We know that if a standard number of hours or number of tasks is prescribed for all trainees, the outcome of training will be individuals who perform at considerably variable levels. A more desirable scenario would be to have all trained individuals performing at least at some benchmark level of performance. This benchmark should be established from objectively assessed performance of experts performing the tasks the trainees will use to train. This would establish the performance criterion required for trainees to be considered proficient. The term *proficient* is used here deliberately to avoid confusion with competency. The *Oxford English Dictionary*, 2nd edition, defines proficient as “advanced in the acquirement of some kind of skill: skilled, adept, expert,” while competence is defined as “sufficiency of qualification; capacity to deal adequately with a subject”. As you can see, the former term is more appropriately applied to the acquisition of a particular technical skill, while the latter carries with it a much broader connotation as to the ability to deal with all aspects of a subject. Since we know that technical skill is only a component of overall surgical competence, we feel that proficiency should be used when discussing skills training.

When setting the proficiency criterion, the “experts” used to set the standard should not all be performing in the top 1% or 5%; rather, they should reflect a representative sample of the proficient population. If the proficiency criterion level is set too high, trainees will never reach it, and if set too low, an inferior skills set will be produced. Ideally, national or international benchmarks for simulator performance proficiency should be set. This would give trainees objectively established goals they would have to reach each level before progressing to the next level. In personal conversations with senior members of the American Board of Surgery, it has been made clear that they find this concept of proficiency-based progression very attractive as an equitable training strategy. While national or international proficiency criterion levels may be some way off, proficiency criterion can be set locally in each training program or hospital. The Yale VR to OR study has shown the power of this approach.<sup>4</sup> The whole point of training is not simply to improve performance but also to make it more consistent. Indeed, performing well consistently is emerging as one of the key indicators of training success.<sup>11,17,18</sup>

The ultimate goal of any training method is to produce surgeons who have the cognitive and psychomotor tools to competently perform procedures. However, the very defini-

tion of what constitutes “competent” is currently very illusive. This is reflected in the variability of recommendations of different international bodies before competence is assumed for various invasive procedures. For flexible gastrointestinal endoscopy, recommendations for minimum case numbers required for competency range from as few as 50 to as many as 300 procedures.<sup>19</sup> In laparoscopic cholecystectomy, the range is between 10 and 50.<sup>20</sup> What this variability illustrates is that the number of cases performed is an exceedingly poor and subjective standard for skills competency. Objective skills assessment and training to objectively determined proficiency criteria through the use of simulation will allow for a more objective definition of “competency” in the future. This also will require a wide array of simulation devices to reflect the broad range of surgical training, although some skills have a definite commonality.

## Metrics and Feedback

To provide the trainee with objective (and proximate) feedback requires the trainer to be able to objectively assess performance through the use of metrics. The formulation of metrics requires breaking down a task into its essential components (task deconstruction) and then tightly defining what differentiates optimal from suboptimal performance. Unfortunately, this aspect of simulation has been given all too little attention by the simulation industry. Drawing on the example from the MIS community, almost all of the VR simulators use time as a metric. Unfortunately, time analyzed as an independent variable is at best a crude and at worst a dangerous metric. For example, in the laparoscopic environment being able to tie an intracorporeal knot quickly gives no indication of the quality of the knot. A poorly tied knot can obviously lead to a multitude of complications. There are only a few reports in the literature that use objective end product analysis<sup>21</sup> due to the difficulty in acquiring this type of information. For example, Satava and Fried<sup>22</sup> have reported the metrics for the entire endoscopic sinus surgery procedure in *Otolaryngology Clinics of North America*.

There is no magic solution to the issue of metrics, and it is almost certain that good metrics will have to be procedure specific. For example, time may not be the most crucial metric for MIS simulations, but for radiographically guided procedures in interventional radiology or cardiology, time and resultant radiation exposure are very critical. The Imperial College laparoscopic group led by Sir Ara Darzi, has been researching economy of hand movement for number of years with an electromagnetic tracking system they have developed (ICSAD).<sup>23</sup> What they have found is that senior or very experienced surgeons have a smoother instrument path trajectory in comparison to less experienced surgeons. The elegance of this approach is that the system can be used to assess open as well as MIS skills. Other groups<sup>17,18</sup> have been using different metrics such as performance variability and

errors as a key indicator of skill level. Senior or experienced surgeons perform well consistently.

The most valuable metrics that a simulation can provide are on errors. The whole point of training is to improve performance, make performance consistent, and reduce errors. One of the major values of simulators is that they allow the trainee to make mistakes in a consequence-free environment before they ever perform that procedure on a patient. The errors that each simulator identifies and provides remediation for will certainly be procedure specific, and the absence of error metrics should cause trainers to question the value of the simulator as a training device. Well-defined errors in simulation allow trainees to experience the operating environment with risks such as bleeding, without jeopardizing a patient. Thus, trainees can learn what they have done wrong and not to make the same mistakes in vivo when operating on patients in the future. Learning is optimized when feedback is proximate to when the error is committed. If simulators are to be used for high-stakes assessment such as credentialing or certification, then the metrics for that simulator must be shown to meet the same psychometric standards of validation as any other psychometric test.

### VR to OR

Although simulation in MIS has received the majority of attention in the medical field, simulation has been successfully used by the military for many years. At one end of this spectrum is the flight simulators, which cost millions of dollars and are used for selection and training of pilots. At the other end are part-task or manikin trainers used to train medical personnel for emergencies unique to military conflict. The problem for the military is how to train these skills to a high-enough level of proficiency that they can be used effectively during armed conflict. This same problem has confronted the medical community since first considering adoption of simulation as a tool for training and evaluation. Most of the studies reporting on the effectiveness of simulation have used poorly defined qualitative, global measures with no clear objective measurement of outcomes. It is not surprising that many potential users of simulation are skeptical about their usefulness, given the quality of the majority of existing evidence.<sup>3</sup>

What is required to convince potential users are well-controlled studies with defined objectives that demonstrate clear cause-and-effect relationships. Experimental control allows these relationships to be clearly elucidated. This rigorous methodology has existed in disciplines such as psychology for about a century. Generally applied in animal studies, the neurosciences, and experimental psychology, rigorous experimental control is applied to the behavior (cognitive or psychomotor) to be studied, allowing causal relationships to become clear. Assessment criteria are clearly and unambiguously defined, end points defined with clear training strate-

gies using the “laws of behavior” as a guide, which in turn have been demonstrated and validated in a century of empirical work. Only one study, the VR to OR study with the MIST-VR from the Yale group, has used this rigorous empirical approach with considerable success with respect to the answers generated and the confidence the surgical community has in these answers.<sup>4</sup> Subjects in this study were trained until they were consistently performing at the same level as experts (ie, trained to previously mentioned proficiency criterion level). Their performance was then evaluated by expert assessors who were also trained until they could reliably assess performance with a high degree of agreement. The results clearly and unambiguously demonstrated the power of simulation as a training tool. This type of work needs to be repeated if a still-somewhat-skeptical surgical community is to be convinced of the power of simulation as a training tool.

### Simulation Integrated in MIS

Simulation when integrated into a well-structured curriculum has the potential to be a very powerful training and assessment tool when properly applied. Inappropriate application of simulation will lead the user to the (erroneous) belief that simulation does not work. So how should simulators be appropriately applied to training and evaluation? Individuals in training should receive introductory instruction or education to set their training goal in context. At the same time, they could be acquiring basic laparoscopic skills training on a skills generalization focused VR trainer. They would practice their skills on simulated tasks incorporating the principles of shaping and fading, with tasks becoming progressively more difficult. Trainees would not progress to the next task until they met objectively defined performance criterion consistently. These proficiency criterion scores will have been defined using the objectively assessed performance levels of expert MIS surgeons working in the same field either locally, nationally, or ideally internationally. Knowledge and psychomotor performance would both be objectively assessed, and the trainee would only progress to further advanced training or indeed into the OR when they had achieved the predefined proficiency criterion levels.

What this would mean for the use of simulators is that much of the basic psychomotor skill acquisition would be achieved over an extended period of time (interval learning), preferably in the residents' own hospital or training unit. When they reach a certain performance level, they could then attend a short training course designed to teach the specifics of procedures. Currently, these courses have to teach and train individuals with different levels of knowledge and skill, which can be very frustrating for all. If trainees come to these short courses with at least some benchmark proficiency criterion level, much more could be achieved. These courses should really be “finishing schools” where important aspects of surgical technique are taught. This should also be the



forum where master surgeons impart wisdom to less-experienced colleagues. Training time in surgery is in short supply, particularly for residents as mandated by the ACGME. Optimal return on training-time investment is essential if standards are to be maintained.

### Implications for Nonsurgical Training

While the focus of this writing is the improvement of training in surgical environments, the conclusions are relevant to nonsurgical, medical training challenges. Cognitive aspects of learning are as important for the first-responder health care provider as for the surgical resident. Skill acquisition and formulation of metrics apply to those engaged in learning medical procedures such as starting an IV, inserting a chest tube, or establishing an airway. The development of training strategies that incorporate simulation into an existing curriculum has relevance to enhanced learning of most medical skills. Finally, the establishing of proficiency criterion levels is appropriate to the acquiring of all medical lore.

### The Future

Surgery has been at the forefront of the assessment and integration of simulation into the medical curriculum. The problems that led surgery to consider simulation as a training solution will continue, both in surgery and other interventional disciplines. Advanced technologies such as robotics will radically change how surgery is practiced but will usher in other problems. For example, it does not make economic sense for a junior resident to acquire his or her basic robotic skills on a machine that costs around \$1 million. The only reasonable solution seems to be simulation. Other disciplines such as interventional cardiology currently have no other satisfactory training strategies other than learning on patients. Societal demands for greater accountability in medical performance and professional requirements for uniformity in training, particularly within the European Union and United States of America, are major driving forces. Optimizing patient safety is also a major consideration. There are also economic pressures that demand accountability and cost-effectiveness in training and responsiveness to reduced trainee-patient exposure. These pressures are forcing the interventional professions to mobilize technologies to support the competence and integrity of the profession. As these advanced technologies emerge, it is important that the medical profession as a whole work to understand and harness them. Other medical disciplines would do well to look to the MIS surgery community for lessons learned, opportunities missed, and current achievements in the application and integration of

simulation technology for improved patient care in the 21st century.

### REFERENCES

1. Satava RM. Virtual reality surgical simulator: the first steps. *Surg Endosc.* 1993;7:203–205.
2. Healy GB. The college should be instrumental in adapting simulators to education. *Bull Am Coll Surg.* 2002;11:10–12.
3. Champion H, Gallagher A. Simulation in surgery: a good idea whose time has come. *Br J Surg.* 2003;90:767–768.
4. Seymour N, Gallagher A, Roman S, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg.* 2002;236:458–464.
5. Pearson A, Gallagher A, Rosser J, et al. Evaluation of structured and quantitative training methods for teaching intracorporeal knot tying. *Surg Endosc.* 2002;16:130–137.
6. Hall J, Ellis C, Hamdorf J. Surgeons and cognitive processes. *Br J Surg.* 2003;16:1746–1752.
7. Eysenck M, Keane M. *Cognitive Psychology: A Student Handbook.* Erlbaum: Hove; 1995.
8. Broadbent D. Selective and Control Processes. *Cognition.* 1981;10:53–58.
9. Ritter E, McClusky D, Gallagher A, et al. Perceptual, visuo-spatial, and psychomotor ability correlates with duration of learning curve on a virtual reality flexible endoscopy simulator. Poster at 2004 Southeastern Surgical Congress, 2004.
10. McClusky D, Ritter E, Gallagher A, et al. Correlation between perceptual, visuo-spatial, and psychomotor aptitude to duration of training required to reach performance goals on the MIST-VR surgical simulator. *Am Surg.* 2004. In press.
11. Gallagher A, Satava R. Virtual reality as a metric for the assessment of laparoscopic psychomotor skills: learning curves and reliability measures. *Surg Endosc.* 2002;16:1746–1752.
12. Kazdin A. *Behavior Modification in Applied Settings.* Pacific Grove: Brooks/Cole Publishing Company; 1998.
13. Catania A. *Learning.* 2nd ed. Englewood Cliffs, NJ: Prentice Hall; 1984.
14. Rosser J, Rosser L, Savalgi R. Objective evaluation of a laparoscopic surgical skill program for residents and senior surgeons. *Arch Surg.* 1998;133:657–661.
15. Metalis S. Effects of massed versus distributed practice on acquisition of video game skill. *Percept Motor Skills.* 1985;61:457–458.
16. Louie K, Wilson M. Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron.* 2001; 29:145–156.
17. Ritter E, McClusky D, Gallagher A, et al. Objective psychomotor skills assessment of experienced and novice flexible endoscopists with a virtual reality simulator. *J Gastrointest Surg.* 2003;7:871–878.
18. Gallagher A, Richie K, McClure N, et al. Objective psychomotor skills assessment of experienced, junior, and novice laparoscopists with virtual reality. *World J Surg.* 2001;25:1478–1483.
19. Cass O. Training to competence in gastrointestinal endoscopy: a plea for continuous measuring of objective end points. *Endoscopy.* 1999;9:751–754.
20. Club TSS. The learning curve for laparoscopic cholecystectomy. *Am J Surg.* 1995;170:55–59.
21. Hanna G, Frank T, Cuschieri A. Objective assessment of endoscopic knot quality. *Am J Surg.* 1997;174:410–413.
22. Satava RM, Fried M. A methodology for objective assessment of errors: an example using endoscopic sinus surgery simulator. *Otolaryngol CNA.* 2002;35:1289–1291.
23. Darzi A, Smith S, Taffinder N. Assessing operative skill. *BMJ.* 1999; 318:887.